Naval Research Laboratory
Washington, DC 20375-5000

NRL Report 9179

A Matrix Model for the Linear Feedback Shift Register

W. P. WARDLAW

Identification Systems Branch
Radar Division

July 6, 1989

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	Approved for public release; distribution unlimited.
4. PERFORMING ORGANIZATION REPORT NUMBER(S) NRL Report 9179	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Naval Research Laboratory 6b. OFFICE SYMBOL (If applicable) Code 5350	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code) Washington, DC 20375-5000	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Naval Air Systems Command 8b. OFFICE SYMBOL (If applicable) APC-209	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State, and ZIP Code) Washington, DC 20361-5000	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 64211N; PROJECT NO.; TASK NO. W1253; WORK UNIT ACCESSION NO. DN180-248

11. TITLE (Include Security Classification)
A Matrix Model for the Linear Feedback Shift Register

12. PERSONAL AUTHOR(S)
Wardlaw,* W. P.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM Jun 87 TO Aug 87	1989 July 6	20

16. SUPPLEMENTARY NOTATION
*Affiliation: Department of Mathematics, U.S. Naval Academy, Annapolis, MD 21402

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Linear feedback shift register (LFSR); Secrecy system; Matrix; Random bit stream

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

In this report, a matrix model is used to discover some of the properties of the linear feedback shift register (LFSR) and to consider its application to security systems.

First the hardware and operation of the LFSR is briefly discussed. Then a representation of the LFSR as a finite state device is used to obtain the matrix model for the LFSR. The matrix model is employed to derive a number of known results about the period of an LFSR as well as some new results concerning subperiods of an LFSR.

Cryptographic applications are suggested by the randomness properties of the LFSR bit stream output. The matrix model provides a concise treatment of the cryptanalysis of the simple LFSR system. Some suggestions are made to improve the security of LFSR secrecy systems.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL Emanuel Vegh	22b. TELEPHONE (Include Area Code) (202) 767-3481 22c. OFFICE SYMBOL Code 5350

DD Form 1473, JUN 86 *Previous editions are obsolete.* SECURITY CLASSIFICATION OF THIS PAGE

CONTENTS

A MATRIX MODEL FOR THE LINEAR FEEDBACK SHIFT REGISTER

INTRODUCTION

A linear feedback shift register (LFSR) is a device that produces a long period pseudorandom bit stream (a sequence of zeros and ones) that is determined completely by the settings on a relatively small number of switches and a relatively short initial bit stream. The length of the period of the output is exponentially related to the key length, i.e., the number of switches whose settings determine the output. This suggests the possibility of using the output of an LFSR as an additive to a plain text bit stream to produce an enciphered bit stream. Indeed, such applications of LFSRs have been and still are made.

But it is important to be aware of some dangers involved in the use of LFSRs in cryptographical applications. A simple LFSR system is vulnerable to cryptanalysis based on the possession of plain text of length twice that of the key, even though the period of the LFSR is much longer. This cryptanalysis is discussed later in this report.

Although the author believes Theorem 3 and the related results on subperiods of LFSRs to be new, much of the material in this report is discussed at length in the literature, notably in the excellent book [1] by Solomon W. Golomb. The cryptanalysis of the LFSR is discussed in Ref. 2 (pp. 121-129), and a basic introduction is given in a short appendix to an article by G. J. Simmons, which is reprinted in Ref. 3 (pp. 290-294). However, some discussions in the literature are, in this author's opinion, a bit hard to follow or are flawed by some basic mathematical errors. The motivation for this report is to provide a correct, coherent, and easily understandable treatment of LFSRs based on a matrix model. The matrix model is chosen because it fits in well with the author's area of expertise and because this approach should be accessible to the intended audience of this work.

Following this introduction, the hardware of the LFSR is briefly discussed and its operational performance is stipulated. The device is then represented as a finite state device. The latter is used to introduce the matrix model, which is then employed to investigate the periodicity and randomness properties of the LFSR. This model is also exploited to explore the cryptanalysis of a simple LFSR bit stream secrecy system. The report ends with two naive suggestions for constructing secure systems based on LFSRs. This matter warrants further study.

DESCRIPTION OF THE LFSR

An n-stage linear feedback shift register (LFSR) consists of a sequence of n binary storage devices (flip-flops, memory locations, registers, etc.) labeled $F_1, F_2, \ldots, F_n$ in Fig. 1. Each device stores either a 0 or a 1. Initially, the value a_{i-1} is stored in device F_i for $i = 1, 2, \ldots, n$. At each pulse of a controlling clock, the value in device F_{i+1} is shifted to device F_i, $1 \leq i \leq n - 1$. The

Manuscript approved July 28, 1988.

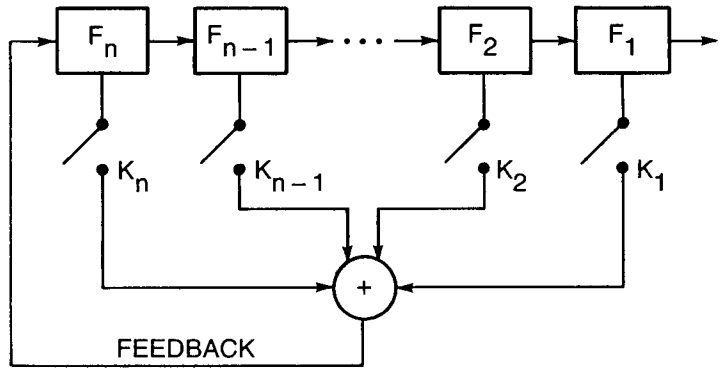


Fig. 1 — Schematic of an n-stage LFSR

new value in F_n is determined by the feedback, which is the sum modulo 2 of the values in those devices F_i for which the switches K_i are closed. Thus, the new value a_{n+k} placed in device F_n at the kth pulse of the clock is given by the nth order recursion

$$a_{n+k} = \sum_{i=0}^{n-1} c_i a_{k+i} \quad \text{(addition modulo 2)}, \tag{1}$$

where each c_i is 0 if switch K_{i+1} is open, or 1 if switch K_{i+1} is closed. The values of the constants c_i comprising the *coefficient vector* $\bar{c} = (c_0, c_1, \ldots, c_{n-1})$ make up an n bit key, and the entries in the *initial state vector* $\bar{a} = \bar{a}(0) = (a_0, a_1, \ldots, a_{n-1})$ make up an n bit initial condition, which together completely determine the output of the shift register. (The two vectors $\bar{c}$ and $\bar{a}$ together can be thought of as a $2n$ bit key for the particular bit stream beginning with $\bar{a}$.)

Equation (1) completely defines the LFSR and its output, the infinite sequence or *bit stream* $A = (a_i) = (a_0, a_1, \ldots)$. This equation is the basis of the remainder of this report.

THE FINITE STATE DEVICE

Instead of viewing the sequence of bits put out by the LFSR, it is useful to consider the *state vectors* $\bar{s} = (s_1, s_2, \ldots, s_n)$ of the LFSR. Here, s_i is the value in the ith binary register F_i. Then we can consider the transition from a given state $\bar{s}$ to the resulting state $\bar{s}'$. This model is called a *finite state device* since its operation is completely described by the transitions $\bar{s} \to \bar{s}'$ among the finitely many state vectors $\bar{s}$. The new state $\bar{s}' = (s_1', s_2', \ldots, s_n')$ is given by $s_i' = s_{i+1}$ for $1 \le i < n$ and $s_n' = \bar{c} \cdot \bar{s}$ · (The latter value is obtained by substituting s_i for a_i in Eq. (1).) Clearly, there are 2^n possible states.

Sometimes it is convenient to represent the state $\bar{s} = (s_1, s_2, \ldots, s_n)$ by the binary notation for the number $n(\bar{s}) = \sum_{i=1}^{n} s_i 2^{n-i}$. For example, (1, 0, 1) and (0, 1, 1) correspond to 101 and 011, respectively. This representation is used in the following examples. The notation $(x.y)$ indicates equation (x) specialized to Example y, as in Eq. (1.1) or Fig. 1.2 below. A similar convention is used to number the figures in the examples.

Example 1. $n = 3, \bar{c} = (1, 0, 1)$.

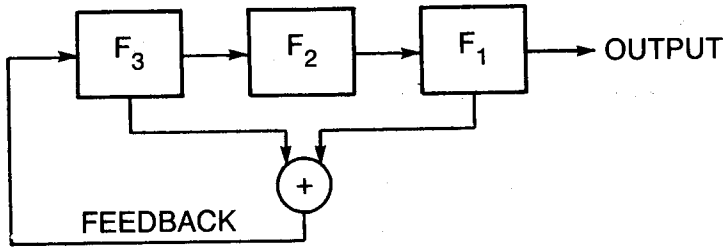
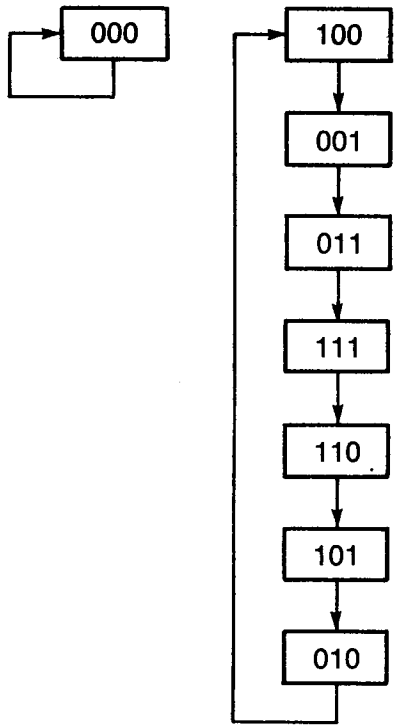


Fig. 1.1 — Schematic of LFSR

$$a_{k+3} = a_k + a_{k+2} \tag{1.1}$$

Finite State Diagram



Bit stream: 1 0 0 1 1 1 0 . 1 0 0 1 1 1 0 . 1 0 0 1 1 1 0. ...

Observe how the binary entries of successive states (in boxes) in Example 1 shift to the left, with a new entry $\bar{c} \cdot \bar{s}$ added on the right. (Other authors use various notations, changing the shift direction and other aspects of the discussion. The interested reader should thoroughly learn one notation; then it will be easy to translate it to any other notation.)

Example 2. $n = 3, \bar{c} = (1, 1, 0)$.

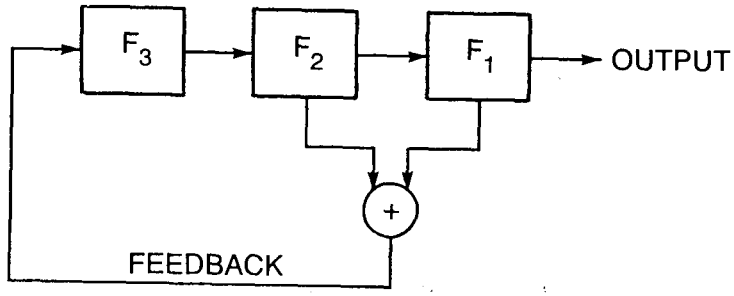
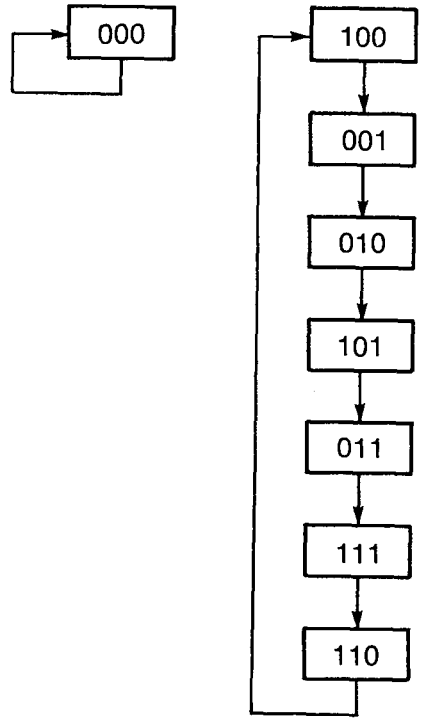


Fig. 1.2 — Schematic of LFSR

$$a_{k+3} = a_k + a_{k+1} \tag{1.2}$$

Finite State Diagram



Bit stream: 1 0 0 1 0 1 1 . 1 0 0 1 0 1 1 ...

Example 3. $n = 3, \bar{c} = (0, 1, 1)$.

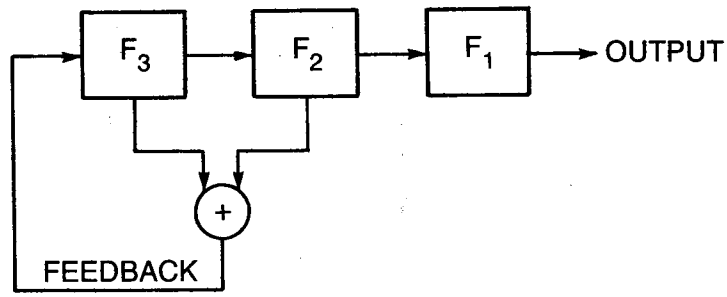
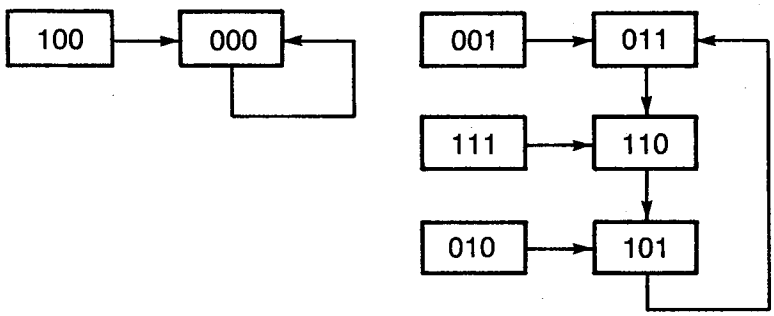


Fig. 1.3 — Schematic of LSFR

Finite State Diagram



Possible bit streams: 1.0.0... , 0.011.011... , 1.110.110... , 0.101.101... , 011.011... .

Example 3 is a *degenerate* three-stage shift register; it is essentially the two-stage LFSR of Example 4, except that the bit stream can begin differently before becoming periodic.

Example 4. $n = 2, \bar{c} = (1, 1)$.

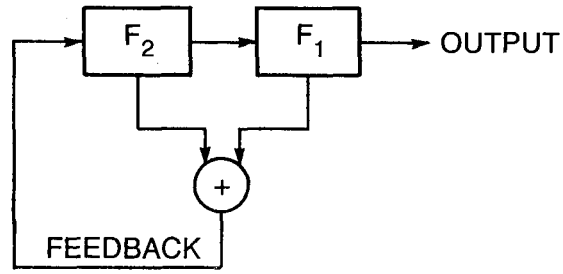
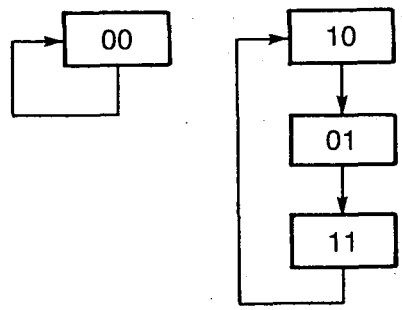


Fig. 1.4 — Schematic of LSFR

$$a_{k+2} = a_k + a_{k+1} \tag{1.4}$$

Finite State Diagram



Bit stream: 101.101.101... .

Note that the bit streams in Examples 3 and 4 are the same except at the beginning, and the nontrivial bit streams have period 3. An LFSR is degenerate whenever there is no feedback from the bit register F_1, or, equivalently, whenever the constant $c_0 = 0$ in Eq. (1).

In all of these examples, the bit stream is periodic. The period turned out to be the number of states in a cyclic chain of states. The state vector of an n-stage LFSR is an n-tuple of zeros and ones, so there are 2^n possible states. Since the zero state $\mathbf{0} = (0, 0, \ldots, 0)$ leads only to itself, it is not taken as an initial state to produce a bit stream. Thus the period cannot exceed $2^n - 1$, the number of nonzero state vectors. A period of $p = 2^n - 1$ will be called *maximum*. The next section gives greater insight on the length of the period of an LFSR.

THE MATRIX MODEL

The transition of the n-stage LFSR from the state $\bar{s} = (s_1, \ldots, s_n)$ to its successor state $\bar{s}' = (s_1', \ldots, s_n')$ is given by the n linear equations

$$\begin{cases} s_i' = s_{i+1} & \text{if } 1 \le i < n, \\ s_n' = c_0 s_1 + \ldots + c_{n-1} s_n. \end{cases} \tag{2}$$

In matrix form,

$$\bar{s}' = \bar{s}M, \tag{3}$$

where $M = (m_{ij})$ is the $n \times n$ matrix with

$$m_{ij} = \begin{cases} 1 & \text{if } i = j + 1, \\ c_{i-1} & \text{if } j = n, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

That is,

$$M = \begin{bmatrix} 0 & 0 & \ldots & 0 & c_0 \\ 1 & 0 & \ldots & 0 & c_1 \\ 0 & 1 & \ldots & 0 & c_2 \\ . & . & \ldots & . & . \\ 0 & 0 & \ldots & 1 & c_{n-1} \end{bmatrix}.$$

If $c_0 = 0$, the matrix M is singular and $\bar{u}M = \mathbf{0}M = \mathbf{0}$ for $\bar{u} = (1, 0, \ldots, 0)$. Thus $\bar{s}' = \bar{s}M = (\bar{s} + \bar{u})M$, and every successor state $\bar{s}'$ has two (or more) precedents, $\bar{s}$ and $\bar{s} + \bar{u}$. This is the degenerate case in which the LFSR is essentially an $(n - 1)$-stage register. (This situation was encountered in Example 3. The three-stage LFSR of Example 3 is essentially the same as the two-stage LFSR of Example 4.)

Henceforth, we will usually assume that $c_0 = 1$ and M is nonsingular. Thus, the mapping of $\bar{s}$ to $\bar{s}' = \bar{s}M$ is a permutation of the 2^n state vectors. The zero vector is sent to itself and therefore initializes a bit stream consisting entirely of zeros.

There are only finitely many $n \times n$ matrices over the two-element field $GF(2) = Z_2 = \{0, 1\}$ of integers modulo 2. Hence, there are integers s and t such that $0 \le s < t$ and $M^s = M^t$. Since M is invertible, this means that $M^h = I$ is the identity matrix for $h = t - s$. Let p be the smallest positive integer such that $M^p = I$; p is called the *period* of M and of the corresponding LFSR.

Now the matrix M defined by Eq. (4) is a *companion matrix* $M = C(m)$ of the polynomial

$$m(x) = x^n - c_{n-1}x^{n-1} - \ldots - c_1 x - c_0, \tag{5}$$

as described in Ref. 4, (p. 190). The characteristic and minimum polynomial of M is $m(x)$. (See Ref. 4, p. 190, Corollary 1 to Theorem 2). We call $m(x)$ the characteristic polynomial of the LFSR corresponding to M. It follows from the definition of the minimum polynomial that $m(M) = 0$, and that $f(M) = 0$ for the polynomial $f(x)$ if and only if $m(x)$ divides $f(x)$. In particular, $M^k - I = 0$ if and only if $m(x)$ divides $x^k - 1$. These remarks prove

Theorem 1. Let M be a nonsingular matrix over a finite field K with minimum polynomial $m(x)$. Then the period p of M is the smallest positive integer such that $m(x)$ divides $x^p - 1$.

The *exponent* of a polynomial $f(x)$ over a field K is defined to be the smallest positive integer k such that $f(x)$ divides $x^k - 1$, or 0, if no such k exists. Theorem 1 shows that the period of a nonsingular matrix over a finite field K is the same as the exponent of its minimum polynomial. The fact that any such matrix has a positive period establishes the fact that if $f(x)$ is a monic polynomial over a finite field K and $f(0) \neq 0$, then f has a positive exponent.

We are interested in the period of the bit stream $a_0, a_1, a_2, \ldots$ of an LFSR, that is, the smallest positive integer q such that $a_{k+q} = a_k$ for all positive integers k. Of course, this depends on the choice of the initial vector $\bar{a} = (a_0, \ldots, a_{n-1})$. Define

$$\bar{a}(k) = (a_k, a_{k+1}, \ldots, a_{k+n-1}). \tag{6}$$

Then

$$\bar{a}(k) = \bar{a}(k-1)M = \bar{a}M^k, \tag{7}$$

where $\bar{a} = \bar{a}(0)$ is the initial vector and k is any positive integer. Thus the bit stream $A = (a_k)$ has period q if and only if q is the smallest positive integer such that $\bar{a}M^q = \bar{a}$. Of course, this does not require that $M^q = I$, but merely that $M^q - I$ be singular. It has already been observed that $q \leq 2^n - 1$; since there are only $2^n - 1$ nonzero n-tuples of zeros and ones, there must be a duplication among the vectors $\bar{a}M^k$ for $k = 0, 1, 2, \ldots, 2^n - 1$.

The zero bit stream has period 1. If a nonzero bit stream has period q, then q is called a *subperiod* of the LFSR. In Examples 1 and 2, $q = 2^3 - 1 = 7$ is the only subperiod, which is also the period p of these LFSRs.

The matrices are

$$M = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} \quad \text{and} \quad M = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad \text{respectively.}$$

In each case, $p = 7 = \min \{k \in \mathbf{N} : M^k = I\}$. This is more easily seen from the minimum polynomials $m(x) = x^3 + x^2 + 1$ and $m(x) = x^3 + x + 1$, respectively. In each case, $m(x)$ divides $x^7 - 1 = (x - 1)(x^3 + x + 1)(x^3 + x^2 + 1)$, but $m(x)$ does not divide $x^k - 1$ for $k < 7$.

In Example 4, $q = 2^2 - 1 = 3$ is the only subperiod, and the period is $p = 3$. In all three of these cases, the period $p = 2^n - 1$ is maximum, and every subperiod is equal to p.

The following three examples illustrate other possibilities for the period and subperiods of an LFSR. In these examples, the state diagrams are abbreviated by omitting the arrows between state vectors and listing the vectors of the form $\bar{a}M^k$ in a column under $\bar{a}$.

Example 5. $n = 3$, $\bar{c} = (1, 0, 0)$, $M = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$.

States:	000	111	001	011	Period: 3.
			010	110	
			100	101	Subperiods: 1, 3, 3.

$$m(x) = x^3 + 1 = x^3 - 1 = (x - 1)(x^2 + x + 1).$$

Example 6. $n = 3$, $\bar{c} = (1, 1, 1)$, $M = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$.

States:	000	111	010	001	Period: 4
			101	011	
				110	Subperiods: 1, 2, 4.
				100	

$$m(x) = x^3 + x^2 + x + 1 = (x - 1)^3.$$

Example 7. $n = 4$, $\bar{c} = (1, 1, 1, 1)$, $M = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$.

States:	0000	0001	0010	0111	Period: 5
		0011	0101	1111	
		0110	1010	1110	Subperiods: 5, 5, 5.
		1100	0100	1101	
		1000	1001	1011	

$$m(x) = x^4 + x^3 + x^2 + x + 1.$$

Observe that in every example given, the period p of the LFSR is also the largest subperiod. This is always the case.

Theorem 2. Let p be the period of the nondegenerate n-stage LFSR with matrix M. Then the bit sequence with initial vector $\bar{a} = (1, 0, \ldots, 0)$ has period p, $n \leq p < 2^n$, and every subperiod q divides p. Moreover, if p is maximum ($p = 2^n - 1$), then p is the only subperiod of the LFSR.

Proof: Let s be the period of the bit stream with initial vector $\bar{a}$. Then s is the smallest positive integer such that $\bar{a}M^s = \bar{a}$, and $\bar{a}(k)M^s = \bar{a}M^{s+k} = \bar{a}M^k = \bar{a}(k)$ for every positive integer k. Thus M^s acts as the identity on the vectors $\bar{a}(1), \bar{a}(2), \ldots, \bar{a}(n)$. But $\bar{a}(1) = \bar{a}M = (0, 0, \ldots, 0, 1) = (a_1, a_2, \ldots, a_n)$ and $\bar{a}(k) = (a_k, a_{k+1}, \ldots, a_{k+n-1})$ begins with $n - k$ zeros followed by $a_n = 1$ in position $n - k + 1$, so the set $\{\bar{a}(1), \bar{a}(2), \ldots, \bar{a}(n)\}$ forms a basis of the vector space K^n of all n-tuples with entries in $K = GF(2)$. Since M^s acts as the identity on a basis, it must be the identity. Hence, s is the smallest positive integer such that $M^s = I$; that is, $s = p$ is the period of M.

Now, p is a subperiod of M, so (as already shown) $p \leq 2^n - 1$. Since $\bar{a}(p + 1) = \bar{a}(1)$, and $\{\bar{a}(1), \bar{a}(2), \ldots, \bar{a}(n)\}$ is independent, it follows that $n \leq p$. (The latter is also a corollary of Theorem 1, since $m(x)$ divides $x^p - 1$ and $m(x)$ has degree n.)

Suppose that q is a subperiod of M. Thus, for some nonzero vector $\bar{v}$, q is the smallest positive integer such that $\bar{v} = \bar{v}M^q$. Let $p = dq + r$ with $0 \leq r < q$. Then $M^p = I$, so $\bar{v} = \bar{v}M^p = \bar{v}M^{dq+r} = \bar{v}(M^q)^d M^r = \bar{v}M^r$. The minimality of q implies that $r = 0$, so q divides p.

Finally, suppose the period p of M is maximum. That is, $p = 2^n - 1$. Then the $2^n - 1$ vectors $\bar{a}(k) = \bar{a}M^k$ for $0 \leq k \leq 2^n - 2$ include all the nonzero vectors in K^n, and each of these vectors has period p. Clearly, then, the only subperiod is $q = p = 2^n - 1$.□

The reader may have noticed from the examples that M has subperiod 1 if and only if $\mathbf{1}M = \mathbf{1}$ for $\mathbf{1} = (1, 1, \ldots, 1)$. This is the case exactly when the coefficient vector $\bar{c}$ has an odd number of ones, and the latter is equivalent to $m(1) = 0$. Thus 1 is a subperiod if and only if $x - 1$ divides $m(x)$. This can be generalized as follows.

Theorem 3. Let $m(x)$ be the characteristic polynomial of an LFSR. Then for any positive integer q, the LFSR has a subperiod q if and only if $gcd(m(x), x^q - 1)$ is not 1 and does not divide $x^k - 1$ for any $k < q$.

Before proving Theorem 3, we apply it to Examples 1 to 6 previously stated. Recall that if $\bar{c} = (c_0, c_1, \ldots, c_{n-1})$, then $m(x) = x^n + c_{n-1}x^{n-1} + \ldots + c_1x + c_0$.

Example 1. $m(x) = x^3 + x^2 + 1$ is irreducible and divides $x^7 - 1 = (x - 1)(x^3 + x + 1)(x^3 + x^2 + 1)$, so the LFSR has period 7. Since $gcd(m(x), x^k - 1) = 1$ for $k < 7$, 7 is the only subperiod, as we already knew from Theorem 2.

Example 2. $m(x) = x^3 + x + 1$ has period and only subperiod 7, exactly as in Example 1.

Example 3. $m(x) = x^3 + x^2 + x = x(x^2 + x + 1)$. ($m(0) = 0$, so the matrix M is singular and the LFSR is degenerate.) Since $gcd(m(x), x^3 - 1) = x^2 + x + 1$ divides neither $x - 1$ nor $x^2 - 1$, the LFSR has subperiod 3. Moreover, $gcd(m(x), x^k - 1) = 1$ unless 3 divides k, so 3 is the only subperiod. Since M is singular, no power of M is equal to I. However, $M^{3+k} = M^k$ for every $k \geq 1$, so we say that M and the LFSR have period 3.

Example 4. $m(x) = x^2 + x + 1$. Since $gcd(m(x), x^3 - 1) = m(x)$ divides neither $x - 1$ nor $x^2 - 1$, 3 is a subperiod. Since $gcd(m(x), x^k - 1) = 1$ unless 3 divides k, in which case $gcd(m(x), x^{3h} - 1) = m(x)$, 3 is the only subperiod. It follows that 3 is also the period. (The latter is also clear because $m(x)$ divides $x^3 - 1$.)

Example 5. $m(x) = x^3 - 1 = (x - 1)(x^2 + x + 1)$. Since $gcd(m(x), x - 1) = x - 1$, the LFSR has a subperiod 1. Since $m(x) = x^3 - 1$, the LFSR has a subperiod and period 3.

Example 6. $m(x) = x^3 + x^2 + x + 1 = (x - 1)^3$. Since $gcd(m(x), x^k - 1) = x^k - 1$ for $k = 1, 2$, and $gcd(m(x), x^4 - 1) = m(x)$, the LFSR has subperiods 1, 2, and 4, and has period 4, since $m(x)$ divides $x^4 - 1$.

The proof of Theorem 3 is facilitated by the following lemmas.

Lemma A. Let A be a square matrix with minimum polynomial $m(x)$, and let $p(x)$ be any polynomial. Then $p(A)$ is nonsingular if and only if $gcd(m(x), p(x)) = 1$.

Proof: Let $d(x) = gcd(m(x), p(x)) = f(x)m(x) + g(x)p(x)$. If $d(x) = 1$, then $I = d(A) = f(A)m(A) + g(A)p(A) = g(A)p(A)$, since $m(A) = 0$. Hence $p(A)$ has inverse $g(A)$, so $p(A)$ is nonsingular. On the other hand, if $d(x)$ has degree ≥ 1, write $m(x) = m_0(x)d(x)$ and $p(x) = p_0(x)d(x)$. Since $m_0(x)$ has lower degree than $m(x)$, $m_0(A) \neq 0$, but $p(A)m_0(A) = p_0(A)m(A) = 0$. Hence, $p(A)$ is singular.□

Lemma B. *(Primary Decomposition Theorem)* Let T be a linear operator on the finite dimensional vector space V over the field K. Let

$$m = p_1^{e_1} p_2^{e_2} \ldots p_r^{e_r}$$

be the factorization of the minimum polynomial m of T into powers of distinct irreducible monic polynomials p_i over K.

Let V_i be the null space of $p_i(T)^{e_i}$, $i = 1, 2, \ldots, r$. Then

(a) $V = V_1 \oplus V_2 \oplus \ldots \oplus V_r$,

(b) each V_i is invariant under T, and

(c) if T_i is the restriction of T to V_i, then the minimum polynomial for T_i is $p_i^{e_i}$.

The proof of this result is given in Ref. 4 (Theorem 12, pp. 180-181).

Lemma C. Let T be a linear operator on the finite dimensional vector space V over the field K with minimum polynomial m, and let p be any polynomial over K. If T_W is the restriction of T to the null space W of $p(T)$, then the minimum polynomial of T_W is $m_W = gcd(m, p)$.

Proof: Since $p(T_W) = 0$, it follows that m_W divides p. Moreover, m_W divides m, since $m(T_W) = 0$. Hence, m_W divides $d = gcd(m, p)$.

Let $d = p_1^{a_1} \ldots p_k^{a_k}$ be a factorization of d into distinct monic polynomials p_i that are irreducible over K. Then

$$p = p_0 p_1^{b_1} \ldots p_k^{b_k} \text{ and } m = m_0 p_1^{e_1} \ldots p_k^{e_k},$$

where p_0 and m_0 are polynomials such that $gcd(p_0, m) = gcd(m_0, p) = 1$. For each $i = 1, 2, \ldots, k$ there is a vector $\bar{v}_i$ in V_i (see Lemma B) such that $\bar{v}_i p_i(T)^{e_i} = \mathbf{0} \neq \bar{v}_i p_i(T)^{e_i - 1}$, since $p_i^{e_i}$ is the minimum polynomial of T restricted to V_i. Now, $a_i = \min(b_i, e_i)$, so for each $i = 1, 2, \ldots, k$, there is a vector $\bar{w}_i$ in V_i such that $\bar{w}_i p_i(T)^{a_i} = \mathbf{0} \neq \bar{w}_i p_i(T)^{a_i - 1}$. (Simply let $\bar{w}_i = \bar{v}_i p_i(T)^{e_i - a_i}$, where $\bar{v}_i$ is the vector found above.) Thus, $\bar{w}_i$ is in W. Now $m_W = p_1^{c_1} \ldots p_k^{c_k}$ with $c_i \leq a_i$, since it divides $d = p_i^{a_1} \ldots p_k^{a_k}$. But $c_j < a_j$ implies $\bar{w}_j m_W(T) = \bar{w}_j p_j(T)^{c_j} (m_W / p_j^{c_j})(T) \neq \mathbf{0}$, since $\bar{w}_j p_j(T)^{c_j}$ is a nonzero vector in V_j and $(m_W / p_j^{c_j})(T)$ acts nonsingularly on V_j. Therefore, each $c_i = a_i$ and $m_W = d$ as claimed.□

Lemma D. Suppose $\bar{v}$ has period q with respect to the matrix A and $\bar{v}A^k = \bar{v}$ for some $k > q$. Then q divides k.

Proof: By definition, q is the smallest positive integer such that $\bar{v}A^q = \bar{v}$. Let $k = qd + r$ with $0 \leq r < q$. Then $\bar{v} = \bar{v}A^k = \bar{v}A^{qd+r} = \bar{v}(A^q)^d A^r = \bar{v}A^r$ implies $r = 0$ by the minimality of q. Therefore, q divides k.□

We are now ready to prove Theorem 3. We apply the lemmas to the matrix M of the LFSR. Lemmas B and C will be applied to the matrix M interpreted as a linear transformation on the vector space K_n of all n-tuples of elements in the field $K = GF(2) = Z_2$ of integers modulo 2.

Proof of Theorem 3: Consider an LFSR with matrix M and characteristic polynomial m. Let q be a positive integer and let W be the null space of $M^q - I$. By Lemma C, $d = gcd(m, x^q - 1)$ is the minimum polynomial of the restriction M_W of M to W. If d divides $x^k - 1$ for $k < q$, then $M_W^k - I = 0$ and $\bar{v}M^k = \bar{v}$ whenever $\bar{v}M^q = \bar{v}$ (i.e., whenever $\bar{v}$ is in W), so q is not a subperiod of M. Hence, if q is a subperiod of M, then d does not divide $x^k - 1$ for any $k < q$. Also, if q is a subperiod, $M^q - I$ is singular, so $d = gcd(m, x^q - 1) \neq 1$ by Lemma A.

On the other hand, suppose $d \neq 1$ and does not divide $x^k - 1$ for any $k < q$. Let $d = p_1^{a_1} \ldots p_r^{a_r}$, where each p_i is a monic irreducible polynomial over K, and let $W = W_1 \oplus \ldots \oplus W_r$ be the primary decomposition of W, as in Lemma B. For each $i = 1, \ldots, r$, let $\bar{w}_i$ in W_i satisfy $\bar{w}_i p_i(M)^{a_i} = \mathbf{0} \neq \bar{w}_i p_i(M)^{a_i - 1}$, and let $\bar{w} = \bar{w}_1 + \ldots + \bar{w}_r$. Now $\bar{w}M^q = \bar{w}$, since $\bar{w}$ is in W. Suppose, if possible, that $\bar{w}$ has period $k < q$. Then k divides q by Lemma D and $x^k - 1$ divides $x^q - 1$, so $gcd(m(x), x^k - 1) = p_1^{b_1} \ldots p_r^{b_r}$ with $b_i \leq a_i$ for $1 \leq i \leq r$. Since d does not divide $x^k - 1$, there is a j such that $b_j < a_j$. But then $\bar{w}(M^k - I)E_j = \bar{w}_j(M^k - I) \neq \mathbf{0}$ (E_j is the projection of W onto W_j, and it commutes with M) since $\bar{w}_j p_j(M)^{b_j} \neq \mathbf{0}$. This contradicts $\bar{w}$ having period k. Therefore $\bar{w}$ has period q, and q is a subperiod of M. This completes the proof of Theorem 3.□

As we have seen, the characteristic polynomial $m(x)$ of an LFSR is sufficient to determine all the periods of the LFSR. It is usually desirable to make the subperiods as large as possible. That is, we want the period p of the LFSR to be the only subperiod. The next corollary shows how to accomplish this goal.

Corollary 4. Let $m(x)$ be the characteristic polynomial of an LFSR with period p. If $m(x)$ is irreducible, or, if p is prime and $m(1) \neq 0$, then p is the only subperiod of the LFSR.

Proof: By Theorem 1, p is the smallest positive integer such that $m(x)$ divides $x^p - 1$. If $m(x)$ is irreducible, it follows that $gcd(m(x), x^k - 1) = 1$ for any $k < p$, and hence Theorem 3 shows that p is the only subperiod. If p is prime and q is a subperiod, q divides p by Theorem 2, so $q = 1$ and $gcd(m(x), x - 1) \neq 1$, again by Theorem 3. But the latter implies $x - 1$ divides $m(x)$ and $m(1) = 0$, contradicting the hypothesis.□

The following example shows that the converse of Corollary 4 is false.

Example 8. $m(x) = (x^4 + x + 1)(x^4 + x^3 + 1)$ divides $x^{15} - 1 = (x - 1)(x^2 + x + 1)(x^4 + x + 1)(x^4 + x^3 + 1)(x^4 + x^3 + x^2 + x + 1)$, but $gcd(m(x), x^k - 1) = 1$ for $k < 15$, so $m(x)$ has period and subperiod 15, but no other subperiods.

It has already been observed that an n-stage LFSR can have period at most $2^n - 1$. Now we investigate how to obtain such maximum period LFSRs.

Lemma 5. If $f(x) \neq x$ is an irreducible polynomial over GF(2) of degree n, then $f(x)$ divides $x^{2^n-1} - 1$.

Proof: The algebraic extension L of $K = GF(2)$ corresponding to $f(x)$ is of degree n, so L has 2^n elements and is the splitting field of the polynomial $x^{2^n} - x = x(x^{2^n-1} - 1)$. Hence, $f(x)$ divides $x^{2^n-1} - 1$. (See Ref. 5, p. 39 Lemma 3.2 and p. 169 Theorem 16.3.)□

Corollary 6. If $m(x)$ has maximum exponent, then $m(x)$ is irreducible.

Proof: As usual, assume $m(x)$ has degree n. If $m(x)$ is reducible, it has an irreducible factor $f(x)$ of positive degree $r < n$. Hence, $f(x)$ divides $x^{2^r-1} - 1$, so $gcd(m, x^k - 1) \neq 1$ for some $k \leq 2^r - 1 < 2^n - 1$, and there is a subperiod $q < 2^n - 1$, by Theorem 3. Thus, Theorem 2 shows that $m(x)$ does not have maximum exponent.□

Now we see that the maximum period $p = 2^n - 1$ can only be achieved when $m(x)$ is irreducible, and in this case, p is also the only subperiod. The latter is of importance, since it guarantees a period of maximum length $2^n - 1$ will be achieved for any choice of a nonzero vector $\bar{a} = \bar{a}(0)$. One problem remains—the irreducibility of $m(x)$ does not guarantee it has maximum exponent. Indeed, the fourth degree polynomial $m(x) = x^4 + x^3 + x^2 + x + 1$ given in Example 7 is irreducible, but it has exponent 5 rather than $2^4 - 1 = 15$. The problem is certainly not insurmountable. Golomb's book in Ref. 1 (p. 40) shows that there are

$$\lambda(n) = \phi(2^n - 1)/n \tag{8}$$

polynomials of degree n that have maximum exponent; (ϕ is the Euler totient function; $\phi(k)$ is the number of positive integers less than k, which are relatively prime to k). Golomb's tables (pp. 62-65 and 97-107) list some of these polynomials of maximum exponent.

However, one can also guarantee that $m(x)$ has maximum exponent as follows.

Corollary 7. If $2^n - 1$ is prime, then each irreducible polynomial $m(x)$ of degree n has maximum exponent.

Proof: Let $r = 2^n - 1$. $m(x)$ divides $x^r - 1$, by Lemma 5, so the companion matrix M of $m(x)$ satisfies $M^r = I$. Hence, M has period p with $n \leq p \leq r$ (by Theorem 2), and p divides r by a standard group theoretical argument (or by Lemma D, using $\bar{v} = (1, 0, \ldots, 0) = \bar{a}$ as vector of

period p given by Theorem 2). Since $r = 2^n - 1$ is prime, $n > 1$. Thus $p > 1$ and p divides the prime r, so $p = r = 2^n - 1$.□

Primes of the form $2^n - 1$ are called *Mersenne primes*. Golomb has a table in Ref. 1 (Table III-1, p. 37) that shows the first 23 Mersenne primes $2^n - 1$ obtained by taking $n = 2, 3, 5, 7, 13, 17, 19, 31, 61, 89, 107, 127, 521, 607, 1279, 2203, 2281, 3217, 4253, 4423, 9689, 9941$, and 11213. Thus, if $m(x)$ is an irreducible polynomial of degree n for any of these values, $m(x)$ has maximum exponent $p = 2^n - 1$.

RANDOMNESS PROPERTIES

A bit stream arising from an n-stage LFSR with maximum period $p = 2^n - 1$ satisfies the following "randomness" properties:

R1. The sequence $a(k)$ for $0 \le k \le 2^n - 2$ contains exactly 2^{n-1} ones and $2^{n-1} - 1$ zeros.

R2. In every period of the bit stream, if $0 < k < n - 1$, there are twice as many runs of k zeros as there are of $k + 1$ zeros, and the number of runs of k ones is the same as the number of runs of k zeros.

R3. The *autocorrelation function* $C(t)$ has two values. Explicitly,

$$pC(t) = \sum_{k=1}^{p} (-1)^{a(k)+a(k+t)} = \begin{cases} p & \text{if } t = 0, \\ -1 & \text{if } 0 < t < p. \end{cases} \tag{9}$$

All of these properties arise from the fact that the $p = 2^n - 1$ vectors $\bar{a}(k)$ with $1 \le k \le p$ contain each nonzero n-tuple of 0's and 1's exactly once. For example, the five-stage LFSR with $m(x) = x^5 + x^2 + 1$ has maximum period $31 = 2^5 - 1$. One period of its bit stream is

$$0000100101100111110001101110101. \tag{10}$$

It has 15 zeros and 16 ones, thus it satisfies R1. The runs of zeros and ones are counted below. (*A* run of L ones is a zero followed by exactly L ones and another zero.) The symbols $N_0(L)$ and $N_1(L)$, respectively, denote the number of runs of zeros and ones, of length L.

L	$N_0(L)$	$N_1(L)$
1	4	4
2	2	2
3	1	1
4	1	0
5	0	1

Thus, the bit stream (10) satisfies property R2. Property R3 also holds for this bit stream. Equation (9) clearly holds for $t = 0$. The reader can check Eq. (9) for $0 < t < p = 31$ by writing the bit stream (10) horizontally and then rewriting it underneath shifted t places to the left and wrapped around to the end. When $t = 5$, one obtains

```
0 0 0 0 1 0 0 1 0 1 1 0 0 1 1 1 1 1 0 0 0 1 1 0 1 1 1 0 1 0 1
0 0 1 0 1 1 0 0 1 1 1 1 1 0 0 0 1 1 0 1 1 1 0 1 0 1 0 0 0 0 1
+ +   + +   +     + +           + + +     +       +   +   + +
```

(The +'s indicate vertical matches.) For each t there will be 15 vertical matches between the two sequences.

S. Golomb proves in Ref. 1 (pp. 43-45) that maximum period LFSR bit streams satisfy the randomness properties R1 through R3. In the interest of brevity, we forego the presentation of his proofs here.

CRYPTANALYSIS OF THE LFSR

The long length $2^n - 1$ of the period of the bit stream compared to the relatively short length $2n$ of the key, as well as the random nature of the bit stream, suggest that the LFSR bit stream be used as an additive to plaintext to produce scrambled text. However, note that the linear relationship between the key and the bit stream output makes the LFSR vulnerable to the following cryptanalysis.

Suppose the antagonist can obtain $2n$ bits of ciphertext y_i for $1 \leq i \leq 2n$ and corresponding plaintext x_i for $1 \leq i \leq 2n$. Since $y_i = x_i + a_i$ (addition in $GF(2) = Z_2$), the corresponding bit stream $a_i = y_i + x_i$ $(1 \leq i \leq 2n)$ can be recovered by using addition modulo 2. Thus the vectors

$$\bar{a}(k) = (a_k, a_{k+1}, \ldots, a_{k+n-1})$$

(Eq. (6)) can be constructed for $1 \leq k \leq n + 1$. Thence the $n \times n$ matrices A and B, whose kth rows are $\bar{a}(k)$ and $\bar{a}(k + 1)$, respectively, can be constructed. Recall from Eq. (7) that

$$\bar{a}(k)M = \bar{a}(k + 1), \tag{11}$$

where M is the matrix of the LFSR. Thus, $AM = B$ and

$$M = A^{-1}B \tag{12}$$

can be obtained by inverting the nonsingular matrix A. (The nonsingularity of A follows from the independence of $\{\bar{a}(1), \ldots, \bar{a}(n)\}$, as shown in the proof of Theorem 2.) Then the matrix M can be used to produce the entire bit stream by Eq. (11), thus completing the cryptanalysis.

The above analysis seems to presume that the cryptanalyst had prior knowledge of the number n of stages of the LFSR. However, this need not be the case. The cryptanalyst can determine n as the number of "lengthened" vectors $\bar{a}'(k) = (a_k, a_{k+1}, \ldots, a_{k+s})$ $s \geq n - 1$ in a maximal independent set $\{\bar{a}'(k) : k = 1, 2, \ldots, n\}$. All the cryptanalyst needs is $2n$ or more consecutive bits of the LFSR bit stream.

POSSIBILITIES FOR SECURE SYSTEMS

In this section we discuss some possible ways to overcome the vulnerability of LFSRs to cryptanalysis. The comments here are only naive suggestions to consider. The security of a secrecy system can only be validated by the failure of the concerted efforts of a team of expert cryptanalysts.

One suggestion is to use two (or more) LFSRs of periods p and q and add their bit streams. The resulting bitstream would have period equal to the least common multiple of p and q, or to the product pq, if p and q were chosen with no common factors. Thus one would want an m-stage and an n-stage LFSR with m and n relatively prime, having necessarily relatively prime maximum periods $p = 2^m - 1$ and $q = 2^n - 1$, respectively. The resulting period of the bit stream would be pq.

This is the same order of magnitude as the maximum period $2^{m+n} - 1$ of a single $(n + m)$-stage LFSR using the "same hardware," i.e., $n + m$ registers. Further study is needed to determine if the adding of the two bit streams would destroy the linearity that caused the weakness in the single LFSR.

Variations of the above scheme could also be used. For example, the output of one LFSR could be added to the feedback instead of to the output of the other LFSR. Clearly, more investigation is needed in these matters.

CONCLUSION

The matrix model of the LFSR provides a powerful tool for analyzing the behavior of the LFSR. For cryptographic applications, one makes the period q of the bitstream long in comparison to the keylength $2n$. This is best achieved by choosing n so that $2^n - 1$ is a Mersenne prime and choosing the characteristic polynomial $m(x)$ of the LFSR to be irreducible. The result is an LFSR that has the period of every nonzero bitstream equal to the maximal period $p = 2^n - 1$ of the LFSR.

Even when optimized as described above, it can be dangerous to depend on the security provided by simple LFSR systems. The matrix model provides a straightforward method of cryptanalysis. However, secure secrecy systems can probably be designed by using LFSRs in more sophisticated ways.

ACKNOWLEDGMENTS

The author acknowledges the many helpful conversations with Walton Bishop, Anthony Gaglione, Allen Miller, Bruce Richter, and Emanuel Vegh, all of which contributed to the writing of this report. The references suggested by Dr. Vegh were especially helpful, as were the exchanges with Prof. Richter on some of the matrix results. The author is grateful for the support he received during the summer at the Naval Research Laboratory; without that support he would not have accomplished this work.

REFERENCES

1. S.W. Golomb, *Shift Register Sequences*, Revised Edition (Aegean Park Press, Laguna Hills, CA, 1982.)

2. C.H. Meyer and S. M. Matyas, *Cryptology: A New Dimension in Computer Data Security* (Wiley, New York, 1982).

3. G.J. Simmons, ed., *Secure Communications and Asymmetric Cryptosystems*, AAAS Selected Symposium 69 (Westview Press, Boulder, CO, 1982).

4. K. Hoffman and R. Kunze, *Linear Algebra* (Prentice-Hall, Englewood Cliffs, NJ, 1961).

5. I. Stuart, *Galois Theory* (Chapman and Hill, New York, 1973).